

Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

Chang, Naicheng, Tsai, Yuchin and Hopkinson, Alan (2010) An evaluation of implementing Koha in a Chinese language environment. Program: Electronic Library & Information Systems, 44 (4) . pp. 342-356. ISSN 0033-0337 [Article] (doi:10.1108/00330331011083239)

This version is available at: <https://eprints.mdx.ac.uk/6621/>

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>

An Evaluation of Implementing Koha in a Chinese Language Environment

Naicheng Chang, Yuchin Tsai and Alan Hopkinson

Authors:

Naicheng Chang, Assistant Professor, General Education Center, Tatung University, Taiwan. E-mail: ncchang@mttu.edu.tw
Yuchin Tsai, Assistant Researcher, National Center for High-performance Computing, National Science Council, Taiwan. E-mail: thomas@nchc.org.tw
Alan Hopkinson, Technical Manager (Library Services), Learning Resources, Middlesex University, UK. E-mail: a.hopkinson@mdx.ac.uk

Abstract

Purpose – This paper evaluates issues of different scripts in the same record (in MARC21 and Chinese MARC(CMARC)) and Chinese internal codes (i.e. double byte character set) when implementing Koha. It also discusses successful efforts in promoting the adoption of Koha in Taiwan, particularly, the contributions from Koha-Taiwan.

Design/methodology/approach – A Koha CMARC/MARC21 testbed was set up to discuss issues of multi-scripts and Chinese internal codes when implementing Koha, and to determine to what extent the various features of CMARC which are not present in MARC21, such as linking fields, can be supported in Koha and what is required to implement them if they are not available.

Findings – In contrast to Western characters, Chinese internal codes cause extra work, as special programming for character conversion is required when working with Koha. This situation is commonly seen in Asian countries such as China, Japan, and Korea. This paper recommends implementing CMARC, MARC21 or even any other types of MARC format in Koha with strong commercial-level support.

Practical implications – Koha-Taiwan serves Koha Chinese users around the world. A successful Koha-Taiwan could be a model around the world.

Originality/value – In recent years, the concept of implementing an open-source library management system is coming to the fore. In Taiwan, the local-based commercial library system covers more than one quarter of the library system market in academic libraries because the company provides better tailor-made support, mainly in scripts and multiple internal codes, than do western-based commercial

library systems. Evaluations and conclusions from this paper will be useful to countries where multi-scripts and double byte character set are issues.

Article type: Technical paper

Keywords: Library Management Systems (LMS); Open Source Software (OSS); Chinese MARC (CMARC); Double Byte Character Set

Word length: 5067

1. Background

1.1 MARC and its descendants

The ability to process characters with diacritics is an important aspect of library automation and information retrieval systems in many parts of the world. Computers were developed in the western world (mainly in the UK and the US) where the basic character set of Latin characters without diacritics is predominantly used and there are no non-Latin characters represented. However, one needs to know if *eleve* will retrieve *élève* or not. One needs to be able to distinguish between *eleve* and *élève*. The situation becomes even more complex with non-Latin alphabets. The UNICODE character set has been developed for this purpose.

Library automation software uses a number of standards which are almost universally applied across the library world so every library system should incorporate them. One of these standards, arguably the most important as it allows the sharing of records created in one library around the rest of the world, is MARC (Machine-Readable Cataloguing). Unfortunately, it has various dialects, one of which, UNIMARC, is used around the world. In Taiwan, a sub-dialect of UNIMARC, Chinese MARC (CMARC), is the most widely used machine-readable format among libraries. Another sub-dialect of UNIMARC, China MARC (CNMARC), is the most widely used machine-readable format in mainland China and since it derives from UNIMARC it is closely related though has been developed independently. The most common dialect of all is that developed originally by the US Library of Congress as LC MARC, later known as USMARC and now known as MARC21. All Library Management Systems (LMS) open-source software (OSS) packages are likely to implement this format. CMARC and USMARC/MARC21 are two formats used for Chinese/Japanese/Korean and western materials in Taiwan's library community.

The Chinese MARC format was first published in 1982 based on the structure of UNIMARC. It is the goal of the Chinese MARC Working Group (1981) to always keep aligned with international standardisation. The third edition of CMARC was published in 1989, and was adopted by most LMSs at that time. Even till now, the third edition is still the most widely used MARC format because it is integrated into the data structure of the LMSs which are commercially available in Taiwan. The latest edition is CMARC 4th edition with update 2001. The major change of the updated 2001 version is to replace the Linking Entry Block (4xx) with an equivalent Related Title Block (5xx). The conversion specifications from CMARC to MARC21 were completed later with the aim of bibliographic records exchange with national or international bibliographic networks such as OCLC.

1.2 OSS and KOHA

The LMS environment has changed considerably over the last few years. A recent trend has been to develop OSS LMSs. Libraries turn to OSS solutions mainly because several OSS library management systems, such as Koha, have been considered modern and mature systems that would fulfil libraries' needs such as:

- OSS are more open to customisation to meet the special demands of libraries;
- OSS features emerge from the user community that have contracted or developed and contributed to them so that other libraries can use and benefit from them.
- Libraries using OSS have more support options than those using proprietary software (Breeding, 2007).

Koha was originally developed for use with MARC21 which is used around the English-speaking world. In the following section, we will discuss the implementation of CMARC on Koha to evaluate to what extent the various features of CMARC which are not present in MARC21 such as linking fields (e.g. the methodologies for linking from a record of a serial to its later or earlier titles) can be supported in Koha, and what is required to implement them if they are not available.

The Koha system was originally developed for the Horowhenua Library Trust in New Zealand by Katipo Communications in 1999. Koha was implemented at the Nelsonville Library (Athens County, Ohio) late in 2003. According to Breeding (2008b), Koha ranks as the first full-featured open-source LMS and serves the most number of libraries, mainly public libraries in the US. In the UK, the first Koha implementation experience was in 2008 at the Royal London Homoeopathic Hospital (Bissels, 2008). In Taiwan, there are currently about 10 libraries using Koha, with the majority of them being primary schools in rural areas. The library school at Fu-Jen

Catholic University has incorporated the Koha system into its courses on library automation systems as a tool for course practicals.

1.3 Koha community in Taiwan and Asia

Koha-Taiwan was initiated by Professor Chingchen Mao in September 2005 with the aim of developing, maintaining, and promoting Chinese Koha. By 2010 Koha-Taiwan (2010) was based in a Google Group and had 58 registered members, and also had its own website (<http://koha-tw.org>) maintained by the National Center for High-performance Computing, National Science Council, Taiwan. Starting in 2006, Koha-Taiwan continues to release Chinese Koha initially with version 2.2.9 and more recently with version 3.0.1.

Koha-Taiwan serves Koha Chinese users around the world. It also is the only Koha community that currently has been developed in Asian countries. At the time of writing (May 2010), feedback and messages from all over the world have cumulated to a total of 1198. Concrete technical contributions from Koha-Taiwan include:

- 'chinesization' of the Koha interface which required over 90,000 characters in both traditional Chinese and simplified Chinese;
- consolidating CMARC tags and integrating it with Koha tags;
- developing the Koha Chinese Z39.50 environment;
- upgrading Koha Chinese searching ability;
- simplifying Koha installation process;
- composing Koha Chinese operations manual.

1.4 Library management system marketplace in Taiwan

Taiwan is an independent island off the mainland of China with a population of just over 20 million and an area of 36,000 km². Politically, it suffers a certain level of isolation but, technically and in the sphere of education, it has links with all the major industrialised countries. University libraries purchase appropriate LMS as could be found in the rest of the world according to their budgets.

Breeding (2008b) indicated the universal situation that few libraries see the need to replace their working systems because libraries are expecting to acquire next-generation interfaces which will support libraries to manage growing collections of electronic resources, and at the same time because of the advantages of OSS LMSs which challenge strictly the traditional licensing models. This is also true in the library automation system marketplace in Taiwan; many LMSs used in academic

libraries are not able to support UNICODE to manage multi-lingual bibliographic records. The decision of selecting a satisfactory LMS would therefore focus on whether the LMS is open, scalable, and the degree of support of LMS customisation (Ke, 2008). Figure 1 shows that the local-based commercial library system TOTALS II covers more than one quarter of the library system market in major academic libraries in Taiwan. The key reasons for this are library budgets and customer satisfaction. The local-based library system company provides better tailor-made support, mainly in scripts and multiple internal codes, than do western-based commercial library systems.

Take in Figure 1

Figure1. Automation systems installed in the academic libraries in Taiwan (Breeding, 2008a)

Library Technology Guides

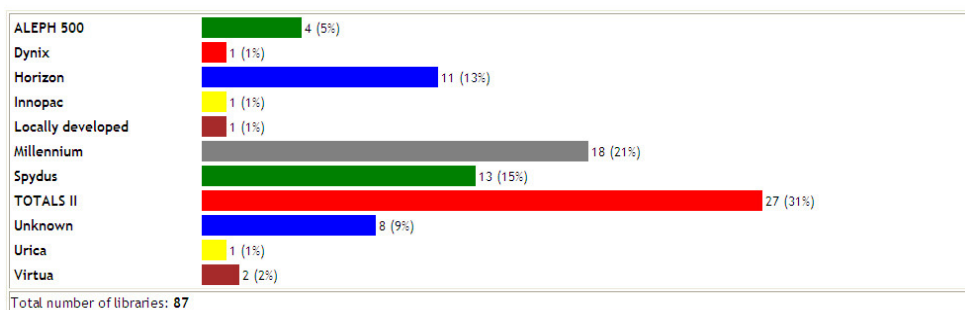
Key resources in the field of Library Automation

Academic Libraries in Taiwan

These graphs indicate what automation systems are used by Academic libraries in based on information recorded in the lib-web-cats database.

Automation Systems Installed

Counting by Library organizations



2. A Koha CMARC/MARC21 testbed

The testbed we are working with is to take Koha as our library system to examine the various features in the MARC formats; that is, CMARC and MARC21. The platform of the testbed is Koha 3.0 with operation system Linux kernel 2.6.24 environment. In the testbed, 100 each of CMARC and MARC21 bibliographic records provided by the National Center for High-performance Computing Library were tested.

The testbed was built to examine mainly three levels:

- the importing of MARC records. To examine if it was properly done when importing MARC records into the Koha system.

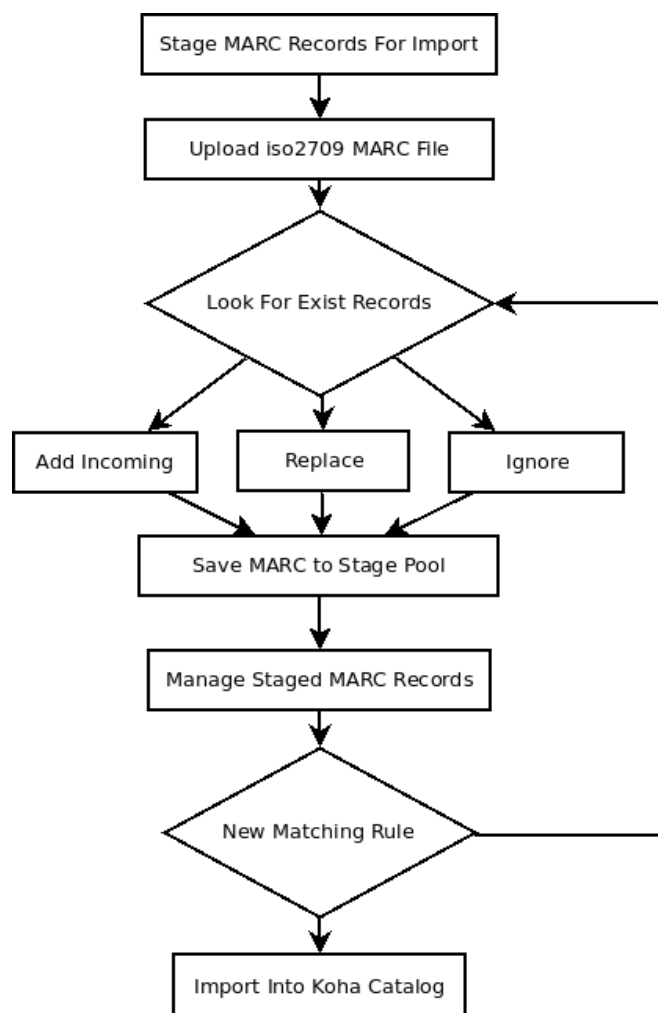
- the integrity of the records after being imported. To examine if bibliographic data could be displayed properly after importing into Koha system.
- to what extent the various features of CMARC which are not present in MARC21 such as linking fields can be supported in Koha and what is required to implement them if they are not available.

2.1 Defining Koha framework

Figure 2 displays the work flow of the testbed, and we will discuss some important steps of the work flow. Before putting Koha to work, we defined the MARC fields and subfields framework using tools provided by the Koha system. Presently, Koha officially provides default MARC21 and UNIMARC templates.

Take in Figure 2

Figure 2. Flowchart of the Koha testbed



It is possible to redefine the cataloguing framework according to the practices of the library. The Koha-Taiwan team has developed a free CMARC template which is currently used in most Koha libraries in Taiwan. A redefining example in CMARC below in Figure 3 shows on the left the default Koha CMARC, and in the example of the right, a subfield \$s (Dynasty) has been redefined in Field 700 (Personal Name - Primary Responsibility), and so is displayed in the catalogue record. The data in subfield \$s remains in Koha even without being redefined.

Take in Figure 3

Figure 3. Redefining Koha framework in CMARC

MARC biblio : 1 (書劍恩仇錄)

With Framework : Default

01278

700 #1 - 人名 -- 主要著者

a 標目主體 金

b 副標目 庸

4 著作方式 著

MARC biblio : 1 (書劍恩仇錄)

With Framework : Default

012789

700 #1 - 人名 -- 主要著者

a 標目主體 金

b 副標目 庸

4 著作方式 著

s 朝代 民國

In the testbed, we defined the testbed to be based on the CMARC 4th edition with update 2001 and MARC21 1999 edition with update 2008.

2.2 Importing MARC records

We encountered no difficulty in importing bibliographic records in the two MARC formats except that the default encoding system in Koha is UNICODE (i.e. UTF-8 here), and in Taiwan, the most accepted internal codes are Chinese Character Code for information Interchange (CCCII) and the 'Big5'. Special programming is required here in order to import the Chinese bibliographic records into Koha and display the data properly. We will discuss the process of special programming later in this Section. An example of CMARC importing data is shown in Figure 4. Clicking the far right end Bib numbers, the system will show the MARC records.

Take in Figure 4

Figure 4. CMARC records

Manage Staged MARC Records › Batch 1

File name	nchc_0513_cmarc.iso
Comments	cmarc(none)
Staged	2010-02-11 09:05:33
Status	staged
Matching rule applied	No matching rule in effect
Action if matching record found	create_new
Action if no match found	create_new
Item processing	always_add

New matching rule	Do not look for matching records ▾
Action if matching record found	Add incoming record ▾
Action if no match found	Add incoming record ▾
Item processing	Always add items ▾

Apply different matching rule

Import into catalog

Page 1 [2](#) [3](#) [4](#)

#	Citation	Status	Match?	Bib
1	書劍恩仇錄 金庸著 (9573204053)	imported	no_match	1
2	未央歌 鹿橋著 (9570500662)	imported	no_match	2
3	風景攝影的藝術 林家祥著 (957532160X)	imported	no_match	3
4	卡內基人緣公式 岱爾·卡內基著 (9579592594)	imported	no_match	4
5	語言與人生 早川S. I. Hayakawa著 (957-32-0372-3)	imported	no_match	5
6	時間簡史 史帝芬·霍金Stephen W. Hawking著 (9575200020)	imported	no_match	6
7	分子的世界 日本分子科學研究振興會編著	imported	no_match	7
8	黃金時代 司馬遷原著 (9573208814)	imported	no_match	8
9	煩惱平息 林清玄著 (9575850491)	imported	no_match	9
10	臺灣熱門旅遊去處 戶外生活雜誌旅遊編輯組編著 (9579476373)	imported	no_match	10

2.3 MARC view and simple view

Bibliographic data can be displayed in the MARC format, in simplified form, or in ISBD format, in both the librarian interface and the OPAC. Before viewing, we defined display formats of fields and subfields. In Figure 5, a CMARC record displays three options of viewing: Normal View, MARC View and ISBD View.

Take in Figure 5

Figure 5. ISBD view of a CMARC record

The screenshot shows the Koha library system interface. At the top, there are navigation links: Circulation, Patrons, Search, and More. A search bar is present with the text "Enter search keywords:" and a "Submit" button. Below the search bar, there are links for "Check Out", "Check In", and "Search the Catalog". The main content area displays the "ISBD Details for 1" record. On the left, there is a sidebar with buttons for "Normal", "MARC", "ISBD", "Items", "Holds", "Checkout history", and "Modification log". The "ISBD" button is selected. The main content area shows the following information: "書劍恩仇錄 [] = Books and Sword, Gratitude and Revenge : / 金庸著 -- 二版", "臺北市 : 遠流, 民81", "2 v. : , 21公分", and "9573204053(平裝) : NT200". At the top right of the main content area, there are buttons for "New", "Edit", "Save", "Add to List", "Print", and "Place Hold".

2.4 Language encoding and conversion

It is essential to have knowledge of the encoding system before working on the Koha system. ISO 2709 (2008) defines the length of fields, yet the length varies due to the characters. For example, Chinese characters coded using the Big5 take 2 bytes, and CCCII take 3 bytes which is different from Western characters, and this causes extra work – i.e. special programming for character conversion, when working on Koha. This situation is commonly seen in other Asian countries, such as China, Japan and Korea.

One of our authors Tsai (2007) developed the conversion programming, and we explain the process as below.

- Step 1: capturing data. This covers capturing data of field tags, subfield tags and coded values according to the character length and control characters as prescribed in the ISO 2709 standard.
- Step 2: converting character encoding. This is done by consulting original text encoding, and applies tools iconv or Perl Encode to convert double byte character set such as Big5 into UNICODE character encoding format UTF-8.

Step 3: redefining. This is done by consulting ISO 2709 to redefine character length of field tags and subfield tags and coded values in the leader.

Step 4: writing conversion result. Saving the new MARC leader, field tags, subfield tags, and coded values into the iso file.

In the testbed, we tested two MARC formats and proved that libraries need to define the MARC framework before importing their data, and the result showed that there is no difficulty for any MARC formats to work on Koha properly.

3. Comparison and evaluation

There are many subtle differences between all the MARC formats, but they are all very closely related to each other because they all act as a carrier for Anglo-American Cataloguing Rules (AACR) and other closely related ISBD-based rules.

As CMARC is derived from UNIMARC, it inherits a more logical tag hierarchy ordered by function. After the coded identifiers/standard numbers and coded data elements (MARC21 begins with codes and follows with identifiers), we find description (ISBD), notes (as in ISBD), linking (which probably follows notes because the traditional output of a link is a note) subject, and then at 700 name access points. Individual tags within the blocks are different. However, these distinctions are in many senses trivial. More importantly, as far as compatibility is concerned, are issues relating to granularity at the subfield level and the methods of linking. As far as implementation of software is concerned, the two features which need to be tested for individual implementation are linking techniques and coded data fields.

Linking is a complex activity. Coded data fields are intricate as far as data entry is concerned. That is because the fields require careful counting of codes to enter and to avoid this being onerous we need a tailor-made entry methodology. Coded fields are important for multi-lingual multi-script functions. Web browsers often attempt to identify scripts and languages in their text and make use of coded data so that the language can more easily be identified. The same holds true for catalogue records. If a system cannot deal with a script, there needs to be transliteration with appropriate information to enable conversion in a system where it is available and to permit information retrieval.

3.1 Comparison

3.1.1 Coded data fields

We tested Coded Data Fields, that is, fixed-length fields, in CMARC on fields 100, 105, 110, 115, 116, 117, 120, 121, 125, 129, 140,141, and in MARC21 on field 007 and 008. We discovered that in the basic Koha system only the default Coded Data Fields 100 and 105 with subfield \$a in CMARC are found, as shown in Table I, yet full subfields in 007 and 008 in MARC21 were defaulted shown in Figure 6 by clicking the icons in the far right end.

Take in Table I

Table I. CMARC default subfields

tagfield	tagsubfield	Data elements
100	a	General Processing Data
105	a	Monograph Coded Data

Take in Figure 6

Figure 6. Cataloguer's screen of 007 and 008 in MARC21

Add MARC Record

Save z39.50 Search Change framework: Default

0 1 2 3 4 5 6 7 8 9

000 - LEADER

001 - CONTROL NUMBER -

003 - CONTROL NUMBER IDENTIFIER

005 - DATE AND TIME OF LATEST TRANSACTION

006 - FIXED-LENGTH DATA ELEMENTS--ADDITIONAL MATERIAL CHARACTERISTICS--GENERAL INFORMATION + -

007 - PHYSICAL DESCRIPTION FIXED FIELD--GENERAL INFORMATION + -

008 - FIXED-LENGTH DATA ELEMENTS--GENERAL INFORMATION

010 ☐ - LIBRARY OF CONGRESS CONTROL NUMBER -

015 ☐ - NATIONAL BIBLIOGRAPHY NUMBER + -

016 ☐ - NATIONAL BIBLIOGRAPHIC AGENCY CONTROL NUMBER + -

020 ☐ - INTERNATIONAL STANDARD BOOK NUMBER + -

022 ☐ - INTERNATIONAL STANDARD SERIAL NUMBER + -

024 ☐ - OTHER STANDARD IDENTIFIER + -

027 ☐ - STANDARD TECHNICAL REPORT NUMBER + -

In the CMARC implementation, the rest of the default subfields of coded data fields are not present by default. This implies that systems librarians need to enable the subfields by redefining the Koha system framework when working with CMARC, whilst all the subfields of Coded Data Fields are available if MARC21 has been used.

For technical support terms of view, this is an advantage of MARC21 when implementing Koha into libraries because not much extra work is needed.

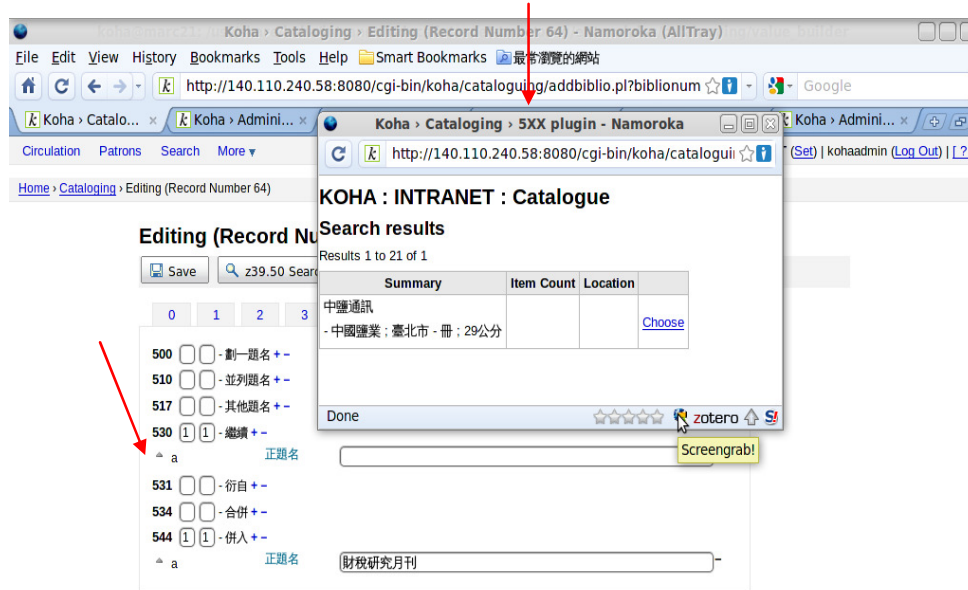
3.1.2 Linking fields

When testing linking fields on the two MARC formats, we discovered that the Koha system does not include by default the linking fields required by any of the two MARC formats. For example, there does not exist the 5xx series of fields in the CMARC default set of fields (in the latest edition of CMARC, the 4xx fields are replaced with the 5xx for the purpose of linking methodology). Users must enable these fields through redefining the framework. It is surprising that these fields are not included but it may be that institutions using Koha do not need linking fields if displaying library catalogue records in an OPAC is their aim. Another possibility is that libraries which do not have many serials or do not use Koha for their serials would find linking fields unnecessary. This situation also is found in some countries that use UNIMARC and do not use linking fields. We believe there are alternative techniques for these to be operational, that is, more programming work would be needed in developing linking mechanism in Koha for any of the MARC formats.

The function of linking fields in Koha is realised mainly through 'Value-builder'. This function is designed to connect instantly relevant bibliographic data together. Librarians could therefore select appropriate data into the library system. Taking the example of CMARC, define firstly `cmarc_field_5xx.pl` to subfield 505\$a (Title), then the Value-builder function will link subfield 505\$a to one of these three subfields: 200\$a or 225\$a or 500\$a. The Koha system also allows users to redefine the linking framework according to an institution's practices if the default function does not meet the institution's need. Furthermore, in CMARC, field 530(Continues), 534(Absorbed) and its subfields also execute Value-builder. Below is an example of CMARC Value-build of bibliographic record 64 in the Koha system. In Figure 7, a `cmarc_field_5xx.pl` was defined to 530\$a (Title). In Figure 8, we see the change history of Title of record 64 through connecting the old title subfield 530\$a 中鹽通訊 to the new title in MARC bibliographic record 64 (鹽業通訊). But then later, it was merged into 544\$a 財稅研究月刊.

Take in Figure 7

Figure 7. cmarc_field_5xx.pl was defined to 530\$a (Title)



Take in Figure 8

Figure 8



3.2 Evaluation

As discussed earlier, CMARC has a more logical tag hierarchy ordered by function than MARC21, yet from a library systems point of view, MARC21 stands in a better position to be applied in the Koha system than CMARC. This is because in MARC21, the most important coded data are designed in 007 and 008, and subfields in the two fields are all defined, while in CMARC, the coded data are spread into more than 10 fields, and more than 90% of their subfields are not defined. Perhaps this explains the reason why the Koha system has grown steadily in the US market more than in any other area in the world.

From their own point of view, institutions which have sufficient technical support might not opt to apply an open-source LMS, but for institutions that do not have sufficient technical staff, and wish to use an open-source LMS, MARC21 could be the first option when considering Koha system because not much extra work is needed when working on coded data fields as discussed earlier in the Section. As to linking fields, although all linking fields are not found in the Koha default subfield, whether institutions can still redefine these accordingly depends on cataloguing policy as well as an institution's technical support arrangements.

The US Koha market has demonstrated a rather successful business model in LibLime, a division of the PTFS (Progressive Technology Federal Systems) company which, according to its website (<http://www.liblime.com/about>) "is the library community's most trusted provider of open-source solutions". LibLime is a company which supports Koha in the same way that traditional LMS companies support their own products. Therefore, we can conclude that implementing CMARC, MARC21, UNIMARC or even any other types of MARC formats in Koha along with strong commercial-level-support is our recommended solution.

3.3 Issue of double byte character set

The double byte character set is commonly known as internal code specification in Asian countries, for instance, Big5 and CCCII in Taiwan, GB 18030-2000 in China, s-jis in Japan and ksc-5601 in Korea. There are around 54,000 CCCII codes, and around 13,000 Big5 codes. Although the library community in Taiwan uses Big5 and CCCII for the Chinese bibliographic materials, the two codes have their own problems. Big5 is popularly implemented in the personal computer and industry sectors, yet it has insufficient characters for libraries to use in terms of maintaining their bibliographic data or patron records. Libraries add characters due to the shortage of Big5 without being compatible with what other institutions are doing, and this causes a severe problem, that is, data cannot be exchanged among libraries because of

the characters the libraries have developed themselves. As to CCCII, mostly its users are found in the library sector rather than in the industry sectors. CCCII meets libraries' need in large quantity of characters to maintain their bibliographic data, yet the major flaw is that the CCCII development team have never built up a good dialogue with LMS commercial companies, LMS users, and application software companies. Thus, not all the LMSs support CCCII due to the small market. According to Mao and Hsu (2006), among the 77 National Bibliographic Information Network (NBINet) member libraries in Taiwan, 38 of them use CCCII, 32 libraries use Big5, and only 7 libraries use UNICODE. It is obvious that UNICODE is a good solution to this chaos.

Converting data from Big5 to UNICODE is expected to have no problems since UNICODE is likely to include all the characters in Big5. To convert from CCCII to UNICODE requires more effort because the CCCII code set has the feature that many codes are mapped to the same identical character. In 2004, the National Central Library in Taiwan hosted an unofficial UNICODE Workgroup with the purpose of seeking a solution to the issues of multiple internal codes (Unicode Workgroup, 2006). The Workgroup covers 16 library members from sectors of higher education, research institutions, and LMS commercial companies. The Workgroup was formed with the mission to build up and maintain code mapping tables as a standard for data conversion from CCCII to UNICODE. Currently, the mapping tables have finalised two-way mapping tables including more than 50,764 mapping sets from CCCII to Unicode and more than 46,000 sets from UNICODE to CCCII. These mapping tables are expected to cover almost all characters that are frequently used. This achievement gives the library community in Taiwan a good start towards a future UNICODE environment.

4. Discussion and conclusion

Koha is a mature integrated library system with good merits. Koha provides default MARC21 and UNIMARC templates. This implies that Koha is designed to be used for MARC21 or UNIMARC. However there is no special provision for formats using multi-scripts like CMARC, JAPAN/MARC or Korean MARC which need special programming. For countries with a lower information technology development, enormous library system technical work is needed which is quite challenging and requires a certain level of computer expertise which is not often found in libraries in many developing countries.

Koha has a more successful implementation experience in terms of scale of adoption and commercial support in the US than in the Europe and Asia. Therefore, implementation of Koha remains a substantial growth market in the US where MARC21 is the main format in use. In the UK, although some libraries now use MARC21, some libraries still follow UKMARC which is more closely related to UNIMARC in the granularity of its fields, which probably is the main reason why Koha is only just starting there. If one wished to implement UKMARC in Koha, there would be a huge effort required to incorporate UKMARC fields and subfields, and it is unlikely that any UK library would wish to change their system now without converting to MARC21 at the same time. Since 2010 there has been a PTFS office in the UK which could support UK systems and this has been used within the library at the Royal London Homoeopathic Hospital to enhance its Koha implementation (Bissels, 2010). However, academic libraries in the UK are slightly deterred from adopting open source because of a report prepared for the Joint Information Systems Committee (JISC) and the Society of College, National and University Libraries (SCONUL) which stated: “Key technological responses include development of open interfaces within a Service Oriented Architecture and developing Web 2.0 models. However, the implementation of an open source LMS is not yet regarded as beneficial” (Adamson, 2008). In Taiwan, although CMARC is closely intertwined with UNIMARC, the double byte character set issue could be potentially complex for both libraries and LMS/IT commercial companies. Currently, there is almost no commercial software company offering support to Koha users because of the small market. Instead, Koha-Taiwan has been giving full support to institutions which wish to set up Koha systems. This situation has been the case until very recently when a commercial software company called Shou Yang Digital Technology advertised a service to offer support to libraries.

It is our view that elements of a successful Koha support model in Taiwan involve efforts from several aspects. Firstly, the role of LMS/IT commercial companies that market, develop and support Koha community. As discussed in Section 3 the issue of double byte character set, although the Unicode Working Group in Taiwan has completed the mapping tables which cover nearly all characters what are frequently used, an ideal Koha commercial support contract would involve:

- the willingness to absorb the MARC format or bibliographic rules changes in the future;
- deliver efficient levels of services;
- meet a high level of customer satisfaction, particularly in the areas of multiple internal codes and Unicode support.

Secondly, as discussed in Section 1, it is important for the Koha community to grow

Koha as a more competitive LMS system than the proprietary commercial LMS systems in functionality, scalability and performance, particularly the next-generation interface. And this is the rationale for which the authors are currently undertaking a Chinese FRBRized Koha prototype project with Web 2.0 concept. This includes an entity-relationship model based FRBR application function library called LibFRBR which will be used to convert existing records of different types of material into FRBRized structures in Koha, and a new generation Online Public Access Catalogue (OPAC) display pattern and on new methodology of searching/retrieving bibliographic records. The FRBR-based systems will be well suited to Taiwan's Chinese Machine-Readable Cataloguing format and Chinese Cataloguing rules. We will test and implement it in a real Koha library when the prototype is ready, so that the results could contribute fruitfully to the community. Furthermore, we suggest that in order to increase market usage, Koha-Taiwan could start with school, public, and special libraries instead of academic libraries which have less room for Koha to promote. This is because school, public, or special libraries have far more limited budgets than academics, and open source LMS system like Koha could be their only option. Lastly, we suggest that Koha-Taiwan could put more effort in promotion including: 1) host workshops around the island targeting librarians, IT personnel, and LMS/IT commercial companies; 2) set up services advisory centres to support libraries remotely, such as the way in which Koha-Taiwan successfully helped University Library of Belize (2010) to set up the Koha system; 3) set up demonstration website to demonstrate functions of Koha system; 4) set up training courses for potential LMS/IT commercial companies around the island; 5) compose Koha manuals for various kinds such as reference manuals.

The growth of these merits discussed above in Koha system involves the users/volunteers distributions from over the world so that the broader community benefits. We expect more contributions from technology, management, and financial sides and that the Chinese Koha community to work together, then a shift from traditional library systems to Koha systems is highly expected. Furthermore, the difficult situation on double byte character set that has encountered in Taiwan is also very true in countries where double-byte is an issue. A successful Koha-Taiwan could be a model around the world.

References (All URLs were checked May 14th 2010)

- Adamson, V. et al. (2008), *JISC & SCONUL Library Management Systems Study. "An Evaluation and Horizon Scan of the Current Library Management Systems and Related Systems Landscape for UK Higher Education"*. Available at: <http://www.jisc.ac.uk/media/documents/programmes/resourcediscovery/lmsstudy.pdf>
- Bissels, G. (2008), "Implementation of an open source library management system experiences with Koha 3.0 at the Royal London Homoeopathic Hospital", *Program: electronic library and information systems*, Vol.42 No.3, pp.303-314.
- Bissels, G. and Chandler, A. (2010), "Two years on: Koha 3.0 in use at the CAMLIS Library, Royal London Homoeopathic Hospital", *Program: electronic library and information system*, Vol.44, No.3, pp.
- Breeding, M. (2007), "An update on open source ILS", *Computers in Libraries*, Vol.27 No.3, pp.27-29.
- Breeding, M. (2008a), "Analysis of library integrated systems marketplace". Paper presented at 2008 Annual Meeting of University Librarians in Taiwan, May 15, 2008, National Chung Hsing University Library, Taichung. Available at: <http://www.lib.nchu.edu.tw/96leaders/pdf/Analysis.pdf>
- Breeding, M. (2008b), "The viability of open source ILS", *American Society for Information Science and Technology Bulletin*, Vol.35 No.2, pp.20-25.
- Chang, Naicheng and Tsai, Yuchin, (2009), "An evaluation of multi-language/multi-script functions in KOHA". Paper presented at World Library and Information Congress: 75th IFLA General Conference and Council, August 23-27, 2009, Milan, Italy. Available at: <http://www.ifla.org/files/hq/papers/ifla75/135-chang-en.pdf>
- Chinese MARC Working Group, Library Automation Planning Committee, (1981), *Chinese MARC Format for Books*, Library Association of China & National Central Library, Taipei, Taiwan.
- ISO 2709, (2008), *Information and documentation -- Format for information exchange*, International Standardization of Organization, Geneva: ISO.
- Ke, H. (2008), "Selection and conversion of integrated library systems". Paper presented at the 2008 Annual Meeting of University Librarians in Taiwan, May 15, 2008, National Chung Hsing University Library, Taichung. Available at: <http://www.lib.nchu.edu.tw/96leaders/pdf/datamation.pdf>
- Koha-Taiwan, (2010), "Koha-Taiwan at Google Groups". Available at: <http://groups.google.com/group/kohataiwan>
- Koha-tw.org, (2010), available at: <http://trac.koha-tw.org/#KOHA> 台灣

Mao, C. and Hsu, C. (2006), "Chinese MARC (Taiwan) and its bibliographic database". Paper presented at World Library and Information Congress: 72nd IFLA General Conference and Council, August 20-24, 2006, Seoul, Korea, Available at: http://archive.ifla.org/IV/ifla72/papers/77-Mao_Hsu-en.pdf

Tsai, Y. (2007), "Special Programming for Character Conversion",
/trunk/cmarc/marc_to_utf8.pl - Koha Taiwan - Trac.
http://trac.koha-tw.org/browser/koha-tw/cmarc/marc_to_utf8.pl
/trunk/cmarc/new_split.pl - Koha Taiwan - Trac.
http://trac.koha-tw.org/browser/koha-tw/cmarc/new_split.pl

Unicode Workgroup, (2006), available at: <http://unicode.ncl.edu.tw>

University Library of Belize, (2010), available at: <http://www.ublibrary.edu.bz/>

Acknowledgements

Financial support for this research from the National Science Council, Taiwan, under the grant NSC 96-2413-H-036-001 and from Tatung University, Taiwan, under the grant B98-W03-084 is gratefully acknowledged.

An earlier version of part of this paper was presented by Chang, Naicheng and Tsai, Yuchin (2009) at the IFLA Congress, Milan, August 2009, and entitled: An evaluation of multi-language/multi-script functions in KOHA.